

**WHAT IS CLAIMED IS:**

1. A process for the production of a peptide, polypeptide, or protein having a predetermined property, comprising:
  - (a) producing a population of sets of nucleic acid molecules that
  - 5 encode modified forms of a target protein;
  - (b) introducing each set of nucleic acid molecules into host cells and expressing the encoded protein, wherein the host cells are present in an addressable array; and
  - (c) individually screening the sets of encoded proteins to identify
  - 10 one or more proteins that have activity that differs from the target protein, wherein each such protein is designated a hit.
2. The process of claim 1, wherein each set of nucleic acid molecules is individually designed and synthesized.
3. The process of claim 2, wherein each set is deposited at a
- 15 locus in an addressable array.
4. The process of claim 1, wherein each polynucleotide in a set encodes a protein that differs by at least one amino acid from the target protein.
5. The process of claim 1, wherein the array comprises a solid
- 20 support with loci for containing or retaining cells; and each locus contains one set of cells.
6. The process of claim 1, wherein the array comprises a solid support with wells for containing or retaining cells; and each well contains one set of cells.
- 25 7. The process of claim 1, wherein the nucleic acid molecules comprise viral vectors; and the cells are eukaryotic cells that are transduced with the vectors.
8. The process of claim 1, wherein the nucleic acid molecules comprise plasmids and the cells are bacterial cells.

9. The method of claim 1, further comprising:

(d) modifying the nucleic acid molecules that encode the hits, to produce a set of nucleic acid molecules that encode modified hits;

(e) introducing the each set nucleic acids that encode the modified hits into cells; and

(f) individually screening the sets cells that contain the nucleic acid molecules that encode the modified hits to identify one or more cells that encodes a protein that has activity that differs from the target protein and has properties that differ from the original hits, wherein each such protein is designated a lead.

10. The process of claim 9, wherein each set nucleic acid molecules in step (d) is individually designed and synthesized.

11. The method of claim 1, wherein the nucleic acid molecules in step (a) are produced by a method selected from among nucleic acid shuffling, recombination, site-directed or random mutagenesis and *de novo* synthesis.

12. The method of claim 9, wherein the nucleic acid molecules in step (d) are produced by a method selected from among nucleic acid shuffling, recombination, site-directed or random mutagenesis, and *de novo* synthesis.

13. The method of claim 2, wherein the nucleic acid molecules in step (a) are produced by a systematically changing each codon in the target protein to a pre-selected codon.

14. The method of claim 13, wherein the codon is selected from a codon encoding Ala (A), Ser (s), Pro (P) or Gly (G).

15. The method of claim 13, wherein the codon is selected from a codon encoding Arg (R), Asn (N), Asp (D), Cys (C), Gln (Q), Glu (E), His (H), Ile (I), Leu (L), Lys (K), Met (M), Phe (F), Thr (T), Trp (W), Tyr (Y) or Val (V).

16. The method of claim 9, wherein the nucleic acids of step (d) are produced by systematically replacing each codon that is a hit, with a codon encoding the remaining amino acids, to produce nucleic acid molecules each differing by at least one codon and encoding modified hits  
5 to identify leads.

17. The method of claim 9, further comprising:  
recombining the nucleic acid molecules encoding the leads;  
introducing those nucleic acid molecules into cells; and  
screening the cells to identify nucleic acid molecules that encode  
10 optimized leads.

18. The method of claim 17, wherein the recombining is two, three or more up to all of the nucleic acids encoding the leads.

19. The method of claim 17, wherein the recombining is effected by a method selected from among nucleic acid shuffling, recombination,  
15 site-directed or random mutagenesis and *de novo* synthesis.

20. The method of claim 1, wherein the modifications are effected in a selected domain of the target protein.

21. The method of claim 1, wherein the modifications are effected along the full length of the target protein.

22. The method of claim 1, wherein the change in activity is at least about 10%, 20%, 30%, 40% or 50%.

23. The method of claim 1, wherein the change in activity is at least about 75%, 100%, 200%, 500% or 1000%.

24. The method of claim 2, wherein at step (b) the titer of the  
25 viral vectors in each set of cells is assessed.

25. The method of claim 24, wherein titering is effected by real time virus titering, comprising:

(i) incubating the nucleic acid molecules or a vector (biological agent) comprising the nucleic acid molecules at an initial  
30 concentration C, which is the unknown titer, with the host cells at a constant known concentration, D;

- (ii) measuring at successive times, an output signal,  $i$ ;
- (iii) determining the time  $t\beta$ , wherein:

$t\beta$  corresponds to  $i = \beta$ ;

$$\beta_{\min} < \beta < \beta_{\max};$$

- 5  $\beta_{\min}$  and  $\beta_{\max}$  correspond to values of  $i$  at the inflection point of the curve  $i = f(t)$ , for the minimal and maximal values, respectively, of the concentrations of a reference biological agent for which the curve  $t\beta = f(c)$  is predetermined; and

- (iv) determining the initial concentration  $C$ .

- 10 26. The method of claim 24, wherein titering is effected by Tagged Replication and Expression Enhancement, comprising:

(i) incubating with host cells a reporter virus vector with a titering virus of unknown titer, wherein the titering virus increases or decreases the output signal from the reporter virus; and

- 15 (ii) measuring the output signal of the reporter virus and determining the titer of the reporter virus;

(ii) determining the titer of the interfering virus by comparing the titer of the reporter virus in the presence and absence of the interfering virus.

- 20 27. The process of claim 9, wherein the nucleic acid molecules comprise viral vectors; and the cells are eukaryotic cells that are transduced with the vectors.

28. The method of claim 26, wherein at step (f) the titer of the viral vectors in each set of cells is determined.

- 25 29. The method of claim 28, wherein the target protein is a protein involved in viral replication.

- 30 30. The method of claim 1, wherein the performance of the screened proteins is evaluated by a Hill analysis or by fitting the output signal to a curve representative of the interaction of the target protein and a test compound.

31. The method of claim 30, wherein the Hill analysis, comprises:

(a) preparing a sample of each nucleic acid molecule or a plasmid or vector that comprises each nucleic acid molecule (biological agent), wherein each sample is obtained by a serial dilution of the molecules or  
5 vector or plasmid at a concentration R1,

(b) incubating each sample of the dilution obtained in (a) with the host cells (target cells) at a constant concentration R2,

(c) determining a P product from the reaction  $R1 + R2$ , at a t moment, in each the sample; and

10 (d) preparing a theoretical curve H from the experimental points R1 and P, for each biological agent by iterative approximation of parameters of the reaction  $R1 + R2 \rightarrow P$ , at the t moment, in accordance with the equation:

$$P = P_{\max} (\pi R1)^r / (\kappa + (\pi R1)^r) \quad r = 1, \dots, n \quad (2)$$

15 in which:

R1 represents the biological agent concentration in a sample from the scale;

R2 is concentration of target cells (*in vitro* or *in vivo*)

P (output) represents the product from the reaction  $R1 +$   
20  $R2$  at a t moment;

$P_{\max}$  represents the reaction maximal capacity;

$\kappa$  represents, at a constant R2 concentration, the biological system for responding to the biological agent (resistance constant R2);

r represents a dependant coefficient of R1 and corresponds  
25 to the Hill coefficient; and

$\pi$  represents the intrinsic power of the R1 biological agent to induce a response in the biological system (P production at the t moment), and

(e) sorting the  $\kappa$  and  $\pi$  values obtained in (d) for each protein  
30 encoded by the nucleic acid molecules or plasmids or vectors and the cells, and then ranking according to the values thereof.

32. The process of claim 1 that is automated.

33. The process of claim 32 that is computer-controlled.

34. A non-random method for generating proteins with a desired property, comprising:

5 identifying a target protein;

preparing sets of variant nucleic acid molecules that each encode a protein that differs by one amino acid from the target protein;

10 screening and selecting the sets of variant nucleic acid molecules to identify those that encode proteins that have activity that differs by a predetermined amount from the activity of the target protein, thereby identifying proteins that are hits;

identifying the residues in the hit proteins encoded by the variant nucleic acid molecules that differ from the target proteins;

15 preparing further sets of variant nucleic acid molecules in which each codon in the nucleic acid molecule encoding each of the identified residues in each of the hits is replaced with codons encoding each of the remaining 18 amino acids to produce the further sets of variant nucleic acids, wherein in each set differs from each other set by one codon; and screening the further sets of nucleic acid molecules to identify

20 those that encode proteins that have activity that differs from the activity of the hits, thereby identifying nucleic acid molecules that encode leads.

35. The method of claim 34, wherein the replaced amino acid positions comprise a functional domain of the protein.

25 36. The method of claim 34, wherein the positions in the protein in which amino acids are replaced comprise at least about 50% of the amino acids in the protein.

37. The method of claim 34, wherein the positions in the protein in which amino acids are replaced comprise at least 90% of the amino acids in the protein.

1002249-1270  
F0422-542200F

38. The method of claim 34, wherein the positions in the protein in which amino acids are replaced comprise at least 95% of the amino acids in the protein.

39. The method of claim 34, wherein the positions in the protein  
5 in which amino acids are replaced comprise all of the amino acids in the protein.

40. The method of claim 34, wherein each set of nucleic acid molecules is generated, processed and screened separately or in parallel.

41. A method for producing a protein having modified properties,  
10 comprising:

(a) preparing a population of nucleic acid molecules that encode rationally modified proteins;

(b) inserting the population into of expression vectors;

(c) introducing each vector into host cells therefor, and expressing  
15 the modified proteins,

(d) screening each modified protein, and selecting one or more that has (have) a modified property.